Tests of Fit Using Sample Moments

IAIN D. CURRIE
Heriot-Watt University

MICHAEL A. STEPHENS
Simon Fraser University

ABSTRACT. A method for constructing tests of goodness-of-fit using moments and generalized least squares was given by Gurland & Dahiya (1970). First, it is shown that the method gives statistics that can be expressed as a sum of components; expressions are found for the components in the normal, exponential, inverse Gaussian and gamma cases. Second, a result is obtained which shows that in some distributions certain parameters are redundant when a test of fit is required; this result leads to a simplification in the calculation of the test statistic. Third, in some cases the parameter estimates in the least squares procedure exhibit a stability property and conditions are given for this to occur. All three of these properties are interpreted in terms of regression models.

Key words: goodness-of-fit, moments, regression, exponential distribution, normal distribution, gamma distribution, inverse Gaussian distribution.

1. Introduction
Under regularity conditions, the infinite set of moments (when these exist) characterize a distribution; thus, it is appealing to use the first s sample moments, or functions of these, as test statistics for goodness of fit. For example, the statistics $h_1$ and $b_1$ have long been used to test for normality (Fisher, 1930), using the first four sample moments. Gurland & Dahiya (1970) and Dahiya & Gurland (1972), in articles hereafter referred to as GD1 and GD2, developed this approach in a systematic way by comparing functions of sample moments to the corresponding functions of population moments. The method depends on being able to find both a vector function $\zeta$ of length s of the first s population moments and a parameterization $\theta$ of the distribution, where $\theta$ has length $q$, $q<s$, such that $\zeta$ is linear in $\theta$. Informally, we can think of the first $q$ components of $\zeta$ being used to estimate $\theta$, while the remaining $s-q$ components are used to make the test of fit. There is a large degree of choice in the method, for not only must a value of s be decided, but there is also no reason to suppose that $\theta$ and $\zeta$ are unique. We comment on these choices in section 4.

The test statistic of Gurland & Dahiya, $\hat{Q}_n$, has an asymptotic $\chi^2$ distribution, $t=s-q$ (GD1, GD2). The main result of this paper is given in section 2 where we show, by applying regression theory, that $\hat{Q}$ has a decomposition into $t$ components $\hat{C}_i$, each with an asymptotic $\chi^2$ distribution, and such that $\hat{Q} = \hat{Q}_{t-1} + \hat{C}_t$. Thus, as a new sample moment is added, the new test statistic contains the previous statistic plus a new asymptotically independent term, and so the statistic $\hat{Q}$ breaks down into components in a manner similar to the Neyman (1937) statistic, or the EDF statistics (Durbin & Knott (1972), Stephens (1974)); each new component might be expected to test for a new departure from the tested distribution. There is a discussion on the use of such components, as opposed to omnibus tests of fit, in the Durbin-Knott paper; see also Jarque & Bera (1980).

We illustrate the method with applications to the normal, exponential, gamma and inverse Gaussian distributions. In these cases, the low order components coincide with tests of fit.
found using other methods of test construction. For example, when applied to a test of
normality, with \( s = 3 \) or 4, the statistics \( b_1 \) and \( b_2 \) result naturally from the method; these were
originally obtained using cumulants by Fisher (1930). In the case of the exponential distribu-
tion, using \( s = 2 \) moments, with only the scale parameter unknown (so \( q = 1 \)) the test statistic
obtained is equivalent to several other statistics already proposed by various authors, often
from quite different points of view (Currie & Stephens (1986a)). Finally, for the gamma and
inverse Gaussian distributions, the first components can essentially be obtained by plotting the
skewness, as measured by \( \mu_3/\alpha^3 \), against the coefficient of variation \( \alpha/\mu \) (Cox & Oakes (1984),
p. 27).

The calculations leading to the components of \( \hat{c} \) can be lengthy. Fortunately, the result of
section 3 substantially reduces the amount of labour involved in many applications: this result
has an interesting interpretation in the regression context. In section 4, we comment on the
choice of \( \theta \) and \( \zeta \), and it is shown that the choice of \( \theta \) and \( \zeta \) is largely unimportant in the
examples considered. Section 5 is concerned with a stability property connected with the
estimation of \( \theta \), and this result is also interpreted from a regression point of view.

2. The decomposition of the test statistic

As far as is practicable, we use the notation of GD1 and GD2. The symbol \( ' \) will be used to
denote both the transpose of a vector or matrix, and to refer to population or sample moments
(\( \mu_i \) and \( m_i \) respectively) about the origin; in context there should be no confusion.

Suppose the null hypothesis is \( H_0 \); a random sample \( X_1, \ldots, X_s \) comes from the continuous
density \( f(x; \theta) \), where \( \theta = (\theta_1, \ldots, \theta_q) \) is a vector of \( q \) unknown parameters. The test statistic
will be based on an estimate of a vector \( \zeta \) of length \( s \); each component \( \zeta_i \), \( i = 1, \ldots, s \) is a
function of the first \( i \) population moments \( \mu_i, j = 1, \ldots, i \). The GD method depends on
choosing \( \theta \) and \( \zeta \) such that \( \zeta \) is linear in \( \theta_1, \ldots, \theta_q \). We write \( \zeta = W\theta \) where \( W \) is an \( s \times q \) matrix of
known constants. Many \( q \)-dimensional parametric families can be parameterized in terms of
the first \( q \) moments; that is to say, the components of \( \zeta \) can be expressed as functions of the
moments \( \mu_1, \mu_2, \ldots, \mu_q \). For such families, the first \( q \) components of \( \zeta \) can always be chosen to
be \( \theta \), if necessary by reparameterizing; in this case, we can take the first \( q \) rows of \( W \) to be the
identity matrix. This will be seen to arise in the normal, gamma and inverse Gaussian examples
discussed later. In general, we require \( W \) to have rank \( q \). The estimate of \( \zeta \) used in the
construction of \( \hat{c} \) is \( h = (h_1, \ldots, h_s)' \), where \( h_i \) is the same as \( \zeta_i \), but with sample moments
\( m_i = \Sigma X_i/n \) replacing the population moments \( \mu_i \); \( h \) is a consistent estimator of \( \zeta \). To fix
ideas, we give a simple example.

Example. Let \( X_1, \ldots, X_s \) be a random sample from the normal distribution \( N(\mu, \sigma^2) \). Define
\( \theta_1 = \mu, \theta_2 = \log \sigma^2 \). Then, as in GD1 and GD2, take \( \zeta = (\mu, \log \mu, \mu, \log (\mu/3))' \). Here, \( \mu_1, \mu_2, \mu_3 \),
and \( \mu_4 \) are moments about the mean and \( \zeta = W\theta \) where

\[
W = \begin{bmatrix}
1 & 0 \\
0 & 1 \\
0 & 0 \\
0 & 2
\end{bmatrix}
\]

The vector \( h \) is then given by \( h = (m_1, \log m_2, m_3, \log (m_4/3))' \). We return to this example later
in the section.

The test statistic is constructed by considering the regression of \( h \) on the columns of \( W \). We have
asymptotically that the expectation of \( h \) is given by

\[
E(h) = \zeta = W\theta.
\]
The asymptotic covariance matrix of \( h \) is

\[ \Sigma = JGJ' / n \]

where \( G \) is the \( s \times s \) symmetric matrix with entries \( G_{ij} = \mu_i' - \mu_j' \), and \( J \) is the \( s \times s \) Jacobian matrix with entries \( J_{ij} = \frac{\partial \xi_i}{\partial \mu_j} \), \( i, j = 1, \ldots, s \). We would like to perform the regression of \( h \) on \( W \) using generalized least squares with \( \Sigma \) as weight matrix. Unfortunately, in general, \( \Sigma \) will depend on \( \theta \). Let \( \hat{\Sigma} \) be a consistent estimator of \( \Sigma \) and now take \( \hat{\theta} \) as the generalized least squares estimator of \( \theta \) in the regression model

\[ h = W\theta + \varepsilon, \quad E(\varepsilon) = 0, \quad \text{Var}(\varepsilon) = \hat{\Sigma}/n. \tag{2.1} \]

This regression model is to be regarded formally with the estimated value of \( \Sigma \) considered as a known fixed quantity: this is allowable since at this stage we are only concerned with the arithmetic of the residual sum of squares. The Gurland-Dahiya statistic \( \hat{Q} \) is the residual sum of squares from this regression; i.e.,

\[ \hat{Q} = n(h - W\hat{\theta})' \hat{\Sigma}^{-1}(h - W\hat{\theta}) \]

with

\[ \hat{\theta} = (W'\hat{\Sigma}^{-1}W)^{-1}W'\hat{\Sigma}^{-1}h. \tag{2.2} \]

The decomposition of \( \hat{Q} \) is obtained by appealing to updating formulae for the residual sum of squares in a regression. Usually, of course, updating formulae are only of computational interest. In the present case, because the successive components of \( h \) involve successively higher moments \( m_i \), it is natural to consider the increase in the residual when \( h \) is extended by including a function involving the next higher moment. It is well known (see, for example, Seber (1977, p. 366)) that if \( \text{RSS}(n-1) \) is the residual sum of squares from a multiple regression with \( n-1 \) data points \( y_1, \ldots, y_{n-1} \), then if a further data point \( y_n \) is added and \( \text{RSS}(n) \) is the residual sum of squares from the regression with \( n \) points then

\[ \text{RSS}(n) = \text{RSS}(n-1) + z_n'z_n. \]

The vector \( z_n \) is not unique but the sum of squares \( z_n'z_n \) is. It immediately follows that

\[ \text{RSS}(n) = \sum_{i=1}^{n} z_i'z_i \tag{2.3} \]

where the component \( z_i'z_i \) corresponds to the addition of point \( i, i > q \) and \( q \) is the number of terms in the model.

The result (2.3) does not apply immediately to the present case since we do not have \( \text{Var}(\varepsilon) = \sigma^2 I \). However, by using the Cholesky decomposition of \( \hat{\Sigma} \) into \( \hat{T}\hat{T}' \), where \( \hat{T} \) is lower triangular, it is easily shown by using the transformation

\[ \eta = \hat{T}^{-1}\varepsilon, \quad U = \hat{T}^{-1}W, \quad g = \hat{T}^{-1}h. \tag{2.4} \]

that (2.3) holds in this more general case too.

We write the decomposition of \( \hat{Q} \), as

\[ \hat{Q} = \hat{Q}_{n-1} + \hat{C}_t = \sum_{i=1}^{t} \hat{C}_i \]

where \( t = s - q \). Thus, the \( r \)th component of the \( \hat{Q} \) statistic is the unique sum of squares that arises as the increase in the residual sum of squares when the moment vector \( h \) of length \( s-1 \) is augmented to length \( s \), \( s = t + q \).
We now turn to the asymptotic distribution of the \( \hat{\sigma} \). We have that in the standard regression model the components \( z_i, i=q+1, \ldots, n \) given in (2.3) are independent \( \chi_i^2 \) variables. Consider first the model (2.1) but with known variance matrix \( \Sigma/n = TT'/n \), where \( T \) is the lower triangular factorization of \( \Sigma \). Then the components in the transformed model \( g = T^{-1}h \) will also be independent \( \chi_i^2 \) variables. Now \( \hat{T} \) is a consistent estimator of \( T \) and thus the components of \( \hat{Q} \) will also be asymptotically independent and asymptotically \( \chi_i^2 \). This completes the proof of:

**Theorem 1**

There exists a set of components \( \hat{\sigma} \), \( i=1, 2, \ldots \) independent of \( t \) such that \( \hat{Q}_i = \Sigma_i^{-1} \hat{C} \), where \( t = s - q \); asymptotically the \( \hat{C} \) are independent each with a \( \chi_i^2 \) distribution.

**Example.** We continue with our example on the normal distribution. With \( h = (m_1, \log m_2, m_3, \log (m_3/3))' \) then (GD1, equation (4.10))

\[
\Sigma = \begin{bmatrix}
\sigma^2 & 0 & 0 & 0 \\
0 & 2 & 0 & 4 \\
0 & 0 & 6\sigma^2 & 0 \\
0 & 0 & 0 & 32/3
\end{bmatrix}
\]  

(2.5)

The lower triangular factorization of \( \Sigma \) into \( TT' \) gives

\[
T = \begin{bmatrix}
\sigma & 0 & 0 & 0 \\
0 & 2^{1/2} & 0 & 0 \\
0 & 0 & 6^{1/2}\sigma^3 & 0 \\
0 & 0 & 8^{1/2} & (8/3)^{1/2}
\end{bmatrix}
\]  

(2.6)

and

\[
T^{-1} = \begin{bmatrix}
1/\sigma & 0 & 0 & 0 \\
0 & 1/2^{1/2} & 0 & 0 \\
0 & 0 & 1/(6^{1/2}\sigma^3) & 0 \\
0 & 0 & -(3/2)^{1/2} & (3/8)^{1/2}
\end{bmatrix}
\]  

(2.7)

We now replace \( \sigma \) by its maximum likelihood estimate \( \hat{\sigma} = m_1^{1/2} \) and this gives \( \hat{T} \). The transformation (2.4) gives \( g = \hat{T}^{-1}h \) with

\[
g' = [m_1/m_1^{1/2}, (\log m_2)/2^{1/2}, m_3/(6m_3)^{1/2}, (3/8)^{1/2} \log m_3/(3m_3)]
\]  

(2.8)

The matrix of regression coefficients is given in (2.4) by \( U = \hat{T}^{-1}W \) and so

\[
U = \begin{bmatrix}
1/m_1^{1/2} & 0 \\
0 & 1/2^{1/2} \\
0 & 0 \\
0 & 0
\end{bmatrix}
\]  

(2.9)

The residual sum of squares follows easily using

\[
L = U(U'U)^{-1}U' = \begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0
\end{bmatrix}
\]  

(2.10)
a particularly simple form to which we shall return in section 5. This gives \( \hat{Q}_2 \) as

\[
\hat{Q}_2 = \frac{n g_1 + n g_2}{\mu} = \hat{\zeta}_1 + \hat{\zeta}_2
\]

where

\[
\hat{\zeta}_1 = \frac{nm_1}{(6m)}; \quad \hat{\zeta}_2 = (3n/8) \left( \log \left( \frac{m_1}{3m} \right) \right)^2.
\]

We note that \( n g_1 \) and \( n g_2 \) are indeed the first two components of \( \hat{Q} \). While this does not follow from the above algebra, it does follow by constructing the \( \hat{Q} \) statistic from the vector \( h = (m_1', \log m_2, m_3)' \); this gives \( \hat{Q}_1 = n g_1 \). A few comments on this example are in order. Component \( \hat{\zeta}_1 \) is equivalent to \( b_1 = m_1/m_3 \), and \( \hat{\zeta}_2 \) is equivalent to \( b_2 = m_2/m_3 \); \( b_1 \) and \( b_2 \) are the well-known measures of skewness and kurtosis, often proposed for testing normality. An interesting aspect of the method is that it reveals two functions of moments, \( \hat{\zeta}_1^2 \) and \( \hat{\zeta}_2^2 \), which are asymptotically \( N(0, 1) \) and independent, and which, when squared, form the components of the overall statistic \( \hat{Q}_2 \). GD1 gives the final statistic \( \hat{Q}_2 \) (there called \( \hat{Q} \) in equation (4.15)) but not the decomposition.

A final point worth noting in this example arises as follows: the estimate of \( \theta \) used implicitly in the calculation of \( \hat{Q}_2 \) is given by \( \hat{\theta} = (U'U)^{-1}U'g \) which yields \( \hat{\theta} = (m_1', \log m_2)' \). Recall that the generalized least squares estimate required an initial estimate of \( \theta \) to substitute in \( \Sigma \). The initial estimate of \( \sigma^2 \) was \( m_3 \); if we had required an initial estimate of \( \mu \) it would have been \( m_1 \). We see now that the re-estimate of \( \theta \) given by (2.2) does not depend on the data vector in the normal way that regression estimates do, but has remained fixed at \( (m_1', \log m_2) \) We shall see in section 5 that this unusual feature depends on an unlikely connection between the model matrix \( W \) and the variance matrix \( \Sigma \).

3. The short method

We saw in section 2 that the DG method led to a test of normality using

\[
\hat{Q}_2 = (n/6)b_1 + (3n/8) \left( \log \left( \frac{b_2}{3} \right) \right)^2 \tag{3.1}
\]

where \( b_1 = m_1/m_3 \) and \( b_2 = m_2/m_3 \). Informally, we might argue that the value of \( \mu \) contributes nothing to the test of fit; it merely locates the normal distribution. It is then reasonable to ask what test statistic results if the parameter \( \mu \) is removed from \( \zeta \). Thus, we take \( \zeta = (\log \mu, \mu, \log (\mu_3/3))' \) and, correspondingly, \( h = (\log m_2, m_3, \log (m_1/3))' \). The \( \mu \) in \( \theta \) becomes redundant and we may take \( \theta = (\log \mu, \mu) \). Hence, \( \theta \) becomes a scalar and \( W \) becomes a vector \( w \), say, where \( w = (1, 0, 2)' \). The \( \Sigma \) matrix for this problem is the lower right \( 3 \times 3 \) submatrix of (2.5) and the required \( T \) and \( T^{-1} \) are obtained immediately from (2.6) and (2.7). We again replace \( \sigma^2 \) by \( m_1 \). Clearly, the particular \( g = (g_2, g_3, g_4)' \) from (2.8), and \( U \) becomes the vector \( u = (1/2^2, 0, 0)' \). A simple calculation now reveals that \( \hat{Q}_2 \) is again given by (3.1). Thus, the statistic given by the reduced problem is identical to that obtained in section 2; the computational burden is however considerably reduced. This is not an isolated example, and the gamma and inverse Gaussian distributions also exhibit this property. We now show that this property is a consequence of the form of the regression matrix, \( W \). The result is stated and proved in a general regression setting.

Theorem 2.

Let \( h, 3 \times 1 \), be a vector of observations from the regression model

\[
h = W\theta + \epsilon, \quad E(\epsilon) = 0, \quad V(\epsilon) = \Sigma, \tag{3.2}
\]
where $\theta$ is $q \times 1$ and $W$ is rank $q$. Suppose that $W$ has the form

$$W = \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix}$$

where $W_1$ is $p \times p$ and non-singular. Let

$$h_1 = \begin{bmatrix} h_1 \\ h_2 \end{bmatrix}, \quad \theta = \begin{bmatrix} \theta_1 \\ \theta_2 \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}$$

where $h_1$, $\theta$, and $\varepsilon$ are $p \times 1$ and $\Sigma_{11}$ is $p \times p$. Define the regression model

$$h_2 = W_2 \theta_2 + \varepsilon_2, \quad E(\varepsilon_2) = 0, \quad V(\varepsilon_2) = \Sigma_{22}.$$  \hspace{1cm} (3.3)

Then the residual sums of squares from models (3.2) and (3.3) are equal.

**Comment.** This would be a trivial result if $\Sigma$ were block diagonal; i.e. $\Sigma_{11} = 0$. However, it does seem a little unexpected that the residual sum of squares for the model (3.2) should not only not depend on $h_1$, but not even depend on the covariance between $h_1$ and $h_2$.

**Proof.** We assume that $\varepsilon$ is Gaussian since this will allow interpretation of the regression calculations. The joint density of $(h_1, h_2)$ is decomposed into the marginal distribution of $h_2$, and the conditional distribution of $h_1$ given $h_2$. Using standard notation we have

$$h_1 | h_2 \sim \text{N}(\mu, \Sigma_{11}) \sim \text{N}(W_1 \theta_1 + W_2 \theta_2 + \Sigma_{12} \varepsilon_2', (h_2 - W_2 \theta_2). \quad \Sigma_{11} - \Sigma_{12} \Sigma_{22}^{-1} \Sigma_{21}).$$

Since $W_1$ is $p \times p$ and non-singular, then for any $r$, $\theta_2$, we can solve exactly for $\theta_1$. Thus, the conditional part of the likelihood function can be fitted perfectly, whatever value of $\theta_1$ is used. Hence, the residual sum of squares is entirely due to the marginal distribution of $h_2$. We have that $h_2 \sim \text{N}(W_2 \theta_2, \Sigma_{22})$ and so the residual sum of squares for model (3.2) is

$$(h_2 - W_2 \theta_2)' \Sigma_{22}^{-1} (h_2 - W_2 \theta_2).$$

This is clearly the residual sum of squares for model (3.3) and the result is proved. We will apply this result in the next section where we give a number of examples, and discuss the choice of $\theta$ and $\zeta$.

### 4. The choice of $\zeta$ and $\theta$

There are many distributions for which it is not easy to construct the statistic $\hat{Q}$. This can happen either because the moments do not exist, or because they exist in a form which makes it difficult to find the required functions of moments. The obvious example in the first case is the Cauchy distribution, while the beta distribution is just one of many distributions for which it appears difficult to find a $\zeta$. In contrast to the Cauchy and beta distributions, there are some distributions for which there is no difficulty in obtaining a $\zeta$. In fact, if $D$ is a lower triangular matrix of constants and $\zeta$ is a vector of moments giving rise to a statistic $\hat{Q}$, then it is trivial to show that $D \zeta$ gives rise to the same statistic. However, there are choices of $\zeta$ which do produce different statistics. We give a number of examples. Computational details have been omitted.

**Example: Normal.** In the example of section 2 we took $\theta = (\mu, \log(\sigma^2))'$ and $\zeta = (\mu_1, \log\mu_2, \mu_3, \log(\mu_3/\mu))'$ which led to

$$\hat{Q}_2 = nb/6 + n(3/8) \{\log(\mu_3/\mu)\}^2.$$
An alternative approach with the normal distribution is to use cumulants (as in Fisher (1930)).
We can take $\theta=(\mu, \sigma^2)'$ and $\zeta=(\kappa_1, \kappa_2, \kappa_3, \kappa_4)'$ and, with the help of the short method, we find
$$
\hat{Q} = nb_1/6 + n(b_2 - 3)/24.
$$
a form given by Jarque & Bera (1980). A further possibility is provided by using the mixed parameterization (Barndorff-Nielsen, 1978) and taking $\theta=(\mu, 1/\sigma^2)'$. Taking $\zeta$ as $(\mu_1, 1/\mu_2, \mu_3, (3/\mu_4)^{1/2})'$, we find
$$
\hat{Q} = nb_1/6 + \frac{3n}{2} \left( \frac{3}{b_2} - 1 \right)^2.
$$
Generalizations of these results for any $t$ are given in Currie & Stephens (1989).

**Example: Exponential.** There are two natural approaches to the exponential distribution,
$F(x) = 1 - \exp\left(-x/\theta\right)$, $\theta > 0$. The first approach is given in GD1 and GD2 and uses
$$
\zeta = (\mu_1, \mu_1/\mu_2, \mu_3/\mu_4, \ldots, \mu_t/\mu_{t+1})'.
$$
Another approach is to use
$$
\zeta = (\mu_1, \mu_2/\mu_1, \mu_3/\mu_2, \ldots, \mu_t/\mu_{t-1})'.
$$
Both these approaches lead to general expressions for $\hat{Q}$ (Currie & Stephens, 1986b). The first component in each case is equivalent to Greenwood's test of uniformity (Greenwood, 1946), a statistic that makes a habit of turning up in a number of different ways (Currie & Stephens, 1986a).

**Example: Gamma.** We mention two possible approaches. We take the p.d.f. in the form
$f(x) = (\Gamma(\alpha)\beta^\alpha) x^{\alpha-1} \exp\left(-x/\beta\right)$. If, for simplicity, we restrict to the first component, then one choice is $\theta = (\alpha\beta, \beta')$ and $\zeta = (\kappa_1, \kappa_2/\kappa_1, (2\kappa_3/\kappa_2)^{1/2})'$. This gives
$$
\hat{Q}_1 = \frac{n\hat{\alpha}}{2\hat{\beta}(3\hat{\alpha} + 1)(\hat{\alpha} + 1)} \left( \frac{k_3}{k_2} - 2 \frac{k_2}{k_1} \right)^2
$$
(4.1)
where $k_1$ is a sample cumulant and $\hat{\alpha}$ and $\hat{\beta}$ are consistent estimators of $\alpha$ and $\beta$. If we use the mixed parameterization $\theta = (\alpha\beta, \beta')$, then we can take $\zeta = (\kappa_1, \kappa_2/\kappa_1, (2\kappa_3/\kappa_2)^{1/2})'$. We find
$$
\hat{Q}_1 = \frac{8n}{\hat{\alpha}(3\hat{\alpha} + 1)(\hat{\alpha} + 1)} \left( \frac{k_3}{k_2} - \frac{2k_1}{k_1} \right)^2
$$
which has asymptotically the same distribution as (4.1). In both these examples the short cut theorem of section 3 applies. Results for the inverse Gaussian distribution can also be obtained (see Currie & Stephens (1986c)).

These examples allow us to make some comments on the choice of $\theta$ and $\zeta$. It is clear that a desirable property of a test of fit is that it be invariant under changes in the parameterization. The different approaches used for the normal, exponential and gamma distributions give asymptotically equivalent results for the low order components. These low order components will be more important than the higher order components for testing fit, and thus the methods are, from a practical point of view, invariant under changes in $\theta$.

As for the choice of $\zeta$, we argue as follows. The individual components of $\hat{Q}_1$ are used to make the test of fit. For example, in the normal case against a skew alternative, the first component would be used; the test is equivalent to the $b_1$ test. If the alternative was symmetric, but long tailed, then the second component would be used, equivalent to testing with $b_2$. Thus,
the alternative determines which component gives greatest power, and more generally, the alternative will determine the best form for $\zeta$. In the example on the normal distribution, $b_1$ and $b_2$ arise with each form of $\zeta$ considered. In the exponential and gamma cases, the first components are asymptotically equivalent for different $\zeta$. This suggests that the choice of $\zeta$, like $\theta$, is not important, and $\theta$ and $\zeta$ can be chosen on grounds of convenience, although the choice of $\theta$ and $\zeta$ could be made on the grounds of the small sample properties of the components.

5. A stability result

The computation of $\bar{Q}$ is a two stage procedure. A convenient estimate $\hat{\theta}$ of $\theta$ is required in order to estimate the covariance matrix $\Sigma/n$. The estimate of $\Sigma$ is then used in a least squares estimation of $\theta$, giving a second estimate, $\tilde{\theta}$. An interesting feature of the normal and exponential cases discussed above is that $\tilde{\theta}$ turns out to be the same as the preliminary estimate $\hat{\theta}$, whatever the length of $h$. This equality is rather surprising since it means that the least squares estimator $\tilde{\theta}$ remains fixed at the value $\hat{\theta}$, whatever the length of $h$. In an ordinary regression situation such behaviour would be unthinkable, since it implies that the estimator $\tilde{\theta}$ does not depend on most of the data! In the present situation, it can only be explained by a peculiar and very strong condition on the covariance matrix of $h$. The required condition is given in theorem 3 below, where, as in theorem 2, we state and prove the result in a general regression setting.

The equality of $\tilde{\theta}$ and $\hat{\theta}$ has consequences for the computation of $\bar{Q}$. Let $\hat{T}$ be the lower triangular matrix such that $\hat{T}\hat{T}^{-1}=\tilde{\Sigma}$. Then, it is easy to show, using standard formulae for the residual sum of squares in generalized least squares, that $\bar{Q}$ can be computed as

$$\bar{Q} = nh\hat{T}^{-1}(I-L)\hat{T}^{-1}h$$

(5.1)

where

$$L = \hat{T}^{-1}W(W^t\hat{\Sigma}\hat{T}^{-1}W)^{-1}W^t\hat{T}^{-1}$$

(5.2)

In particular cases $L$ has a very simple form. We saw in our introductory example on the normal distribution that $L$ was given by

$$L = \begin{bmatrix} I_2 & 0 \\ 0 & 0 \end{bmatrix}$$

(5.3)

This example can be generalized for the normal and exponential distributions (Currie & Stephens (1989, 1986b)). The relationships between the stability result for the estimation of $\theta$, the condition on $W$ and $\Sigma$, and the computation of $\bar{Q}$, using $L$ are given in:

Theorem 3

Let $h, s \times 1$, be a vector of observations from the regression model

$$h = W\theta + \varepsilon, \quad E(\varepsilon) = 0, \quad V(\varepsilon) = \Sigma$$

(5.4)

where $\theta$ is $q \times 1$ and $W$ is rank $q$. Suppose that $W$ has the form

$$W = \begin{bmatrix} W_1 \\ W_2 \end{bmatrix}$$
where \( W_i \) is \( q \times q \) and non-singular. Let

\[
\begin{bmatrix}
    h_1 \\
    h_2 
\end{bmatrix}
\quad \varepsilon =
\begin{bmatrix}
    \varepsilon_1 \\
    \varepsilon_2 
\end{bmatrix}
\quad \Sigma =
\begin{bmatrix}
    \Sigma_{11} & \Sigma_{12} \\
    \Sigma_{21} & \Sigma_{22} 
\end{bmatrix}
\]

where \( h_1 \) and \( \varepsilon_i \) are \( q \times 1 \) and \( \Sigma_{11} \) is \( q \times q \). Define \( \hat{\theta} = W_i^{-1} h_1 \), and let \( \hat{\theta} \) be the generalized least squares estimate of \( \theta \) in the model (5.4). Let \( \Sigma = TT' \) be the lower triangular factorization of \( \Sigma \)

and define \( L = T^{-1} W (W' \Sigma^{-1} W)^{-1} W' T^{-1} \). Then the following conditions are equivalent:

1. \( L = \begin{bmatrix} I & 0 \\ 0 & 0 \end{bmatrix} \)
2. \( \hat{\theta} = \theta \)
3. \( \Sigma_{11}^2 \Sigma_{11}^{-1} = W_i W_i^{-1} \).

If (1), (2) and (3) hold, then the residual sum of squares is given by

\[
n(\varepsilon_1 - W \hat{\theta}) \Sigma_{11}^{-1} (\varepsilon_1 - W \hat{\theta})
\]

where \( \Sigma_{11} = \Sigma_{11} - \Sigma_{12} \Sigma_{21} \Sigma_{11} \).

**Proof.** The proof is straightforward and is omitted, but the following is instructive. As in theorem 2, we assume that \( \varepsilon \) is Gaussian and use results from regression theory. We require that there be no information on \( \theta \) in \( h_1 \) in addition to that in \( h_i \). This will be the case if

\[
E(h_i | h_i) = W_i \theta + \Sigma_{11} h_i
\]

is free of \( \theta \); i.e. condition (3). The estimate of \( \theta \) is thus computed from \( h_i \), and since \( W_i \) is non-singular, the fixed estimate of \( \theta \) is equal to \( W_i^{-1} h_i \). As in theorem 2, the residual sum of squares comes only from the conditional distribution, which is the final part of the result.

This result poses some interesting theoretical questions. The normal and exponential distributions are the only distributions that we know of for which the stability property holds, for we were unable to find a \( \theta \) and a \( \zeta \) for either the gamma or the inverse Gaussian distributions which led to \( \hat{\theta} = \theta \). Two questions come to mind. Does a parameterization \( \theta \) and a vector of moments \( \zeta \) exist for the gamma or inverse Gaussian distributions for which the stability property holds? A much stronger question is: are the normal and the exponential distributions the only distributions for which the stability result holds?

### 6. Conclusions

The Gurland-Dahia method leads to goodness-of-fit statistics based on sample moments. The present paper discusses some theoretical properties of the technique; in particular, the statistic is shown to have a representation as a sum of asymptotically independent components. Two general remarks are worth making. First, in the examples discussed, the components of \( \hat{Q}_n \), for small \( n \), are shown to be identical to well known tests derived by a wide variety of different methods. Second, we showed in the examples that the low order components are asymptotically independent of the choice of \( \theta \) and \( \zeta \); this can be set against the element of arbitrariness that exists in the choice of \( \theta \) and \( \zeta \). These remarks suggest that the Gurland-Dahia method can profitably be used to produce easily computed tests of goodness-to-fit.

**Acknowledgements**

It is a pleasure to record our thanks for the thorough work of the editor and the referees. Their comments greatly improved the clarity with which we were able to report our results.
References


Received August 1989, in final form February 1990.

Iain D. Currie, Department of Actuarial Mathematics and Statistics, Heriot-Watt University, Riccarton, Currie, Edinburgh EH14 4AS, Scotland.